Understanding Controls on Nitrate Concentration in Groundwater: A Tandem Evolutionary Algorithm Approach

Shreya Chatterjee (shreyachatterjee@ku.edu), University of Kansas; Dr. Erin Seybold, Kansas Geological Survey, University of Kansas

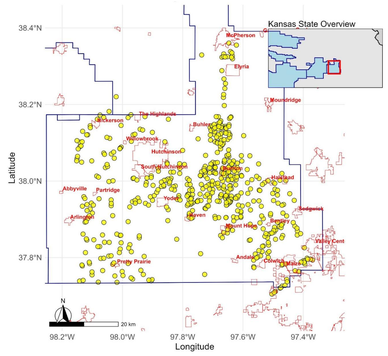
Background

Groundwater contamination by nitrate poses major health and environmental risks.
Previous research has shown that nitrate levels vary dramatically due to complex interactions among land use, soil, aquifer geochemistry, and climate.
Traditional statistical approaches often fail to capture these nonlinear, multivariate interactions.
Machine learning (ML) offers potential for uncovering hidden relationships that drive contamination risk.

Research Questions

- What combination of environmental and anthropogenic factors control nitrate concentration in groundwater and its variability across space and time?
- Can a novel ML approach (TEVA) improve prediction of samples exceeding the EPA Maximum Contaminant Level (10 mg/L)?

Study Region and Data Sources

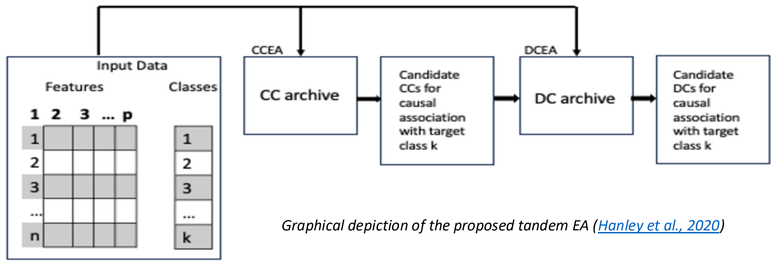


Region: Groundwater Management District 2 (GMD2), Kansas

Map showing the Groundwater Management District 2 (GMD2) in Kansas, highlighting the distribution of monitoring stations (yellow circles).

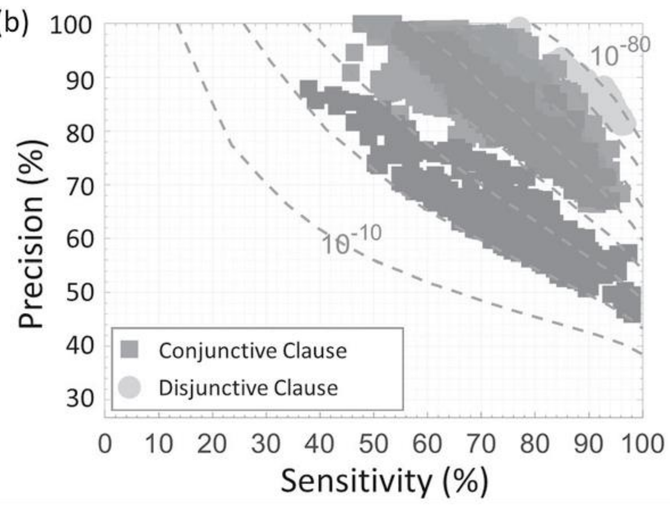
Data Source	Variable Type	Notes
GMD2 Monitoring Network	Nitrate	3,059 samples, 752 wells (1974–2024)
KGS WWC5	Lithology	Depth-resolved sediment composition
NLCD	Land Cover	Cultivated crops, pasture/hay, developed, herbaceous
SSURGO	Soil Properties	Texture, permeability, organic matter
WIZARD	Groundwater Levels	Annual measurements, interpolated
PRISM	Climate	Precipitation, temperature, SPEI
NuGIS	Fertilizer	County-level N application rates
WIMAS	Water Use	Irrigational, municipal, industrial
Census.gov	Demographics	County-level population density

Methodology: Tandem Evolutionary Algorithm (TEVA)



Graphical depiction of the proposed tandem EA (Hanley et al., 2020)

- Goal: Classify groundwater samples as High Nitrate (>10 mg/L) or Low Nitrate (<10 mg/L).
- Input Structure: n×p matrix of predictors
- Factor Selection via Evolutionary Algorithm
- Conjunctive Clauses (CC): Combinations of variables that interact to produce an outcome, joined by "AND".
- Disjunctive Clauses (DC): Combinations of variables demonstrating equifinality, where different variable sets lead to the same outcome, joined by "OR".

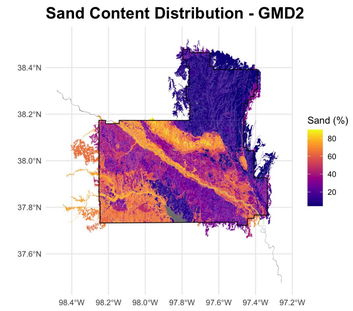


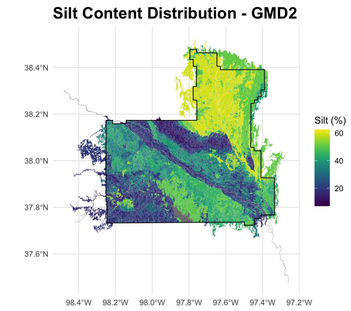
Hypothetical plot of evolved clauses and their precision versus sensitivity (Underwood et al., 2023)

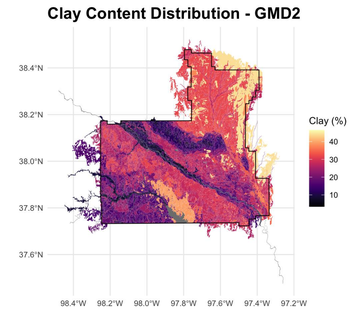
Binary classification (High vs. Low Nitrate)
Metrics:
•Precision = TP / (TP + FP)
•Sensitivity = TP / (TP + FN)
Visual evaluation: Precision–Sensitivity plots (e.g., clauses closer to upper-right corner are strongest)

Current Status & Next Steps

- Assembling and cleaning multi-source dataset
- Calculating predictor metrics within 1 km radius per sample
- Conducting exploratory data analysis (distribution, correlations, spatial trends)

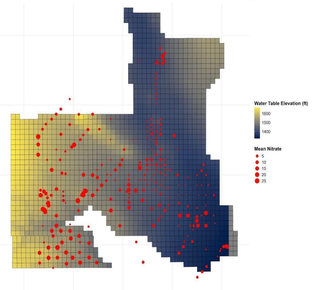






Spatial distribution of soil texture components (clay, sand, and silt) in Groundwater Management District 2 (GMD2), Kansas. The data were extracted from the SSURGO soil database (30-meter spatial resolution).

Water Table Elevation (2022-2024 Avg) – GMD2



The average water table elevation (2022–2024) in GMD2, interpolated at the PLSS section level. The color scale represents elevation, with yellow indicating higher (~1600 ft) and dark blue indicating lower (~1400 ft) elevations. Red circles denote mean nitrate concentrations at monitoring stations, with circle size proportional to nitrate levels

Next Steps

- Implement TEVA classification model
- Evaluate variable interactions and rule structure
- Interpret environmental significance of derived clauses
- Prepare results for publication in 2025 dissertation chapter